# ANALISIS REGRESI

## DENGAN MENGGUNAKAN APLIKASI KOMPUTER STATISTIK SPSS

FRIDAYANA YUDIAATMAJA

# ANALISIS REGRESI DENGAN MENGGUNAKAN APLIKASI KOMPUTER STATISTIK SPSS

# ANALISIS REGRESI DENGAN MENGGUNAKAN APLIKASI KOMPUTER STATISTIK SPSS

Fridayana Yudiaatmaja

Penerbit PT Gramedia Pustaka Utama, Jakarta

# DAFTAR ISI

# PRAKATA

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa karena atas segala rahmat-Nya saya dapat menyelesaikan buku *Analisis Regresi Dengan Menggunakan Aplikasi Komputer Statistik SPSS* (Dilengkapi Dengan Perhitungan Manual) sesuai dengan waktu yang telah direncanakan.

Analisis regresi adalah salah satu analisis yang banyak sekali digunakan oleh mahasiswa maupun dosen di berbagai bidang ilmu pengetahuan untuk mengetahui apakah pengaruh variabel bebas bersifat signifikan terhadap variabel terikatnya serta seberapa besar pengaruhnya. Keunggulan analisis ini adalah kemampuannya dalam melakukan prediksi variabel terikat berdasarkan data-data yang terdapat pada variabel bebas. Analisis regresi dalam perkembangannya juga dijadikan sebagai dasar untuk melakukan analisis jalur. *Structural Equation Modelling* (SEM) merupakan gabungan antara analisis jalur dan analisis faktor yang utamanya digunakan untuk menganalisis variabel-varaibel laten. Pada dasarnya analisis regresi juga merupakan bagian dari SEM. Oleh karena itu penting mendalami analisis ini terutama sekali untuk membantu menjawab masalah-masalah yang timbul di dalam penelitian.

Buku ini disusun berdasarkan telaah pustaka beberapa buku yang berkaitan dengan analisis regresi dan aplikasi

komputer statistik. Dari literatur yang ada, penulis berkeinginan menyinkronisasikan hasil perhitungan manual dengan hasil perhitungan aplikasi komputer statistik SPSS. Hal ini dilakukan agar pembaca dapat mengetahui bagaimana cara mencari angka-angka yang tertera pada output progam SPSS, selain makna angka-angka tersebut. Besar harapan saya buku ini mudah dipahami oleh kalangan akademisi, baik dosen maupun mahasiswa.

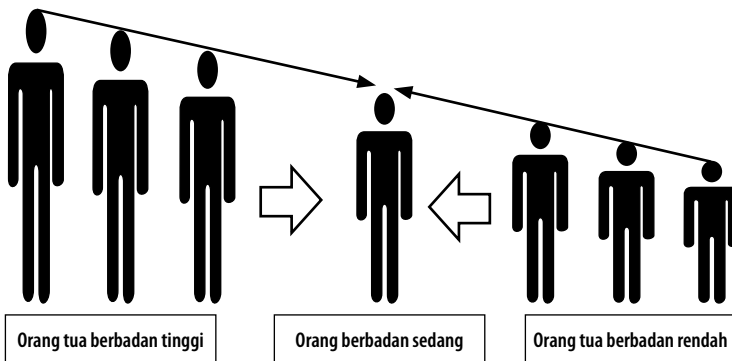
Perlu penulis sampaikan bahwa aplikasi komputer statistik adalah salah satu mata kuliah yang diajarkan di Jurusan Manajemen, Fakultas Ekonomi Universitas Pendidikan Ganesha. Namun mata kuliah tersebut memberi  mahasiswa dasar-dasar analisis data yang menggunakan aplikasi komputer statistik, termasuk analisis regresi, dan analisis-analisis lainnya. Oleh karenanya buku ini tidak mencakup semua materi yang diajarkan mata kuliah tersebut, melainkan sebagai salah satu referensi yang membantu pemahaman mahasiswa mendalami analisis regresi serta sebagai alat bantu analisisnya.

Kepada pembaca, saya sampaikan terima kasih atas kesediaannya menggunakan buku ini dalam mendalami analisis regresi dan alat bantu analisisnya. Akhir kata, segala kritik yang membangun akan saya terima dengan senang hati demi penyempurnaan buku ini.

Singaraja, 12 Februari 2013
Fridayana Yudiaatmaja

# BAB I
# PENDAHULUAN

Analisis regresi adalah metode analisis yang sangat populer di kalangan mahasiswa dan dosen utamanya untuk keperluan penelitian. Pada prinsipnya analisis ini digunakan untuk membuat suatu persamaan yang kelak diharapkan dapat membantu pihak-pihak yang membutuhkan dalam memprediksi nilai variabel terikat dari variabel-variabel bebas di dalam persamaan tersebut. Singkatnya, keunggulan analisis regresi adalah kemampuannya dalam meramalkan atau memprediksi nilai variabel terikatnya.

Istilah regresi ini ternyata sudah dikenal sejak tahun 1886 yang ketika itu Francis Galton menemukan adanya kecenderungan orang tua yang bertubuh tinggi akan memiliki anak dengan tubuh yang tinggi pula dan sebaliknya orang tua yang bertubuh pendek akan memiliki anak yang juga bertubuh pendek (Ghozali, 2005). Hal tersebut didukung oleh Karl Pearson dan A. Lee (Suliyanto, 2011). Di samping itu, Galton juga menemukan bahwa tinggi anak cenderung bergerak ke arah tinggi rata-rata populasi secara keseluruhan yang ditunjukkan oleh gambar yang berada di tengah di dalam ilustrasi berikut ini. Oleh karenanya, arah panah orang yang bertubuh pendek dan orang yang bertubuh tinggi menuju ke tengah yakni orang dengan tinggi badan sedang yang dianggap sebagai tinggi rata-rata populasi.

Ilustrasi Model Regresi Francis Galton (1886), dan Pearson and Lee (1903).
Sumber: Suliyanto (2011)

Namun saat ini analisis regresi umumnya digunakan dengan tujuan mengidentifikasi variabel-variabel bebas mana saja yang signifikan dalam memengaruhi variabel terikat dan seberapa besar perubahan variabel-variabel bebas tersebut dapat memengaruhi perubahan terhadap variabel terikatnya dengan menghitung koefisien variabel-variabel bebasnya. Hal ini sejalan dengan apa yang dikemukakan oleh Sunyoto (2011), bahwa tujuan analisis regresi adalah untuk mengetahui ada tidaknya pengaruh signifikan antara satu atau lebih variabel bebas terhadap variabel terikatnya baik secara parsial atau simultan.

Metode yang umum digunakan untuk menghitung koefisien variabel-variabel bebas yang dihasilkan di dalam persamaan regresi adalah metode *Ordinary Least Square* (OLS). Menurut Gujarati (2006), metode ini paling banyak digunakan karena memiliki sifat teoritis yang kokoh. Metode ini dikenal juga dengan istilah metode klasik di mana persamaan yang dihasilkan akan pas jika jumlah

kuadrat residualnya (*e*) adalah minimum. Dikuadratkan karena jika hanya dijumlahkan saja maka ada kemungkinan nilai residualnya akan menjadi 0. Dengan metode OLS, nilai konstanta dan koefisien suatu persamaan dapat diperoleh dengan meminimumkan persamaan berikut:

$$\sum_{i=1}^{n} e_1{}^2 = \sum_{i=1}^{n} [Y_i - f(X_i)]^2 = minimum$$

*n* adalah jumlah data, *e* adalah residual, *Y* adalah data aktual dan *f*(*X*) adalah fungsi yang dapat menghasilkan data atau nilai prediksi variabel terikat bila diketahui nilai variabel bebas. Lambang *f*(x) ini biasanya juga ditulis dengan Ŷ (Y topi). Lambang tersebut digunakan untuk membedakan data aktual (*Y*) dan data prediksi (Ŷ). Nilai residual (*e*) merupakan selisih antara data aktual dan data prediksinya. Jadi konsepnya adalah semakin kecil jumlah kuadrat residualnya, maka semakin bagus persamaan yang dihasilkan. Sudah tentu banyak sekali persamaan yang bisa dibuat dari suatu data, namun yang terbaik menurut metode OLS adalah persamaan yang residualnya minimum.

Membuat suatu persamaan yang menyatakan hubungan antara variabel terikat dan variabel bebas secara manual relatif mudah jika jumlah variabel yang digunakan sedikit. Namun bukan pekerjaan yang mudah jika jumlah variabelnya banyak. Untuk itu diperlukan alat bantu dalam menghitung sehingga memudahkan kita menghasilkan persamaan yang sesuai dengan data yang ada. Banyak sekali perangkat lunak komputer yang dapat digunakan untuk menghasilkan suatu persamaan, misalnya Microsoft Excel yang menyediakan *Analysis ToolPak-VBA* seperti yang dijelaskan

oleh Atmaja (2009) atau *Statistical Package for the Social Sciences* (SPSS) yang belakangan ini dikenal dengan istilah *Statistical Product and Service Solutions*. Pada kesempatan ini penulis menggunakan SPSS sebagai alat analisis karena aplikasi ini banyak digunakan oleh kalangan mahasiswa dan dosen selain itu secara khusus dibuat sebagai alat uji statistik. SPSS adalah aplikasi komputer statistik yang dibuat oleh IBM, yang tidak hanya digunakan demi kepentingan ilmu sosial,  melainkan juga ilmu lainnya. Versi *trial* atau uji coba perangkat lunak SPSS dapat diunduh melalui www.ibm.com atau www.spss.com.

# BAB II

# ANALISIS REGRESI SEDERHANA

## 2.1 KONSEP DASAR

Yang dimaksud dengan analisis regresi sederhana di sini adalah analisis yang dilakukan terhadap satu variabel terikat dan satu variabel bebas. Biasanya variabel bebas dilambangkan dengan huruf X dan variabel terikat dilambangkan dengan huruf Ŷ. Persamaan yang dihasilkan nantinya mempunyai bentuk Ŷ = a + bX. Jika memiliki sejumlah data aktual X dan Y, maka nilai a dan b dapat dihitung melalui metode OLS.

Di dalam metode OLS, nilai a dan b dapat dihitung dengan meminimalkan fungsi berikut:

$$\sum_{i=1}^{n} e_1{}^2 = \sum_{i=1}^{n} [Y_i - f(X_i)]^2 = minimum$$

Karena *f*(*x*) = Ŷ, maka *f*(*x*)= a + bX, maka persamaan di atas dapat ditulis menjadi:

$$\sum_{i=1}^{n}[Y_i - f(X_i)]^2 = \sum_{i=1}^{n}[Y_i - (a + b\ X_i)]^2 = minimum$$

Untuk meminimalkan suatu fungsi, maka turunan pertama dari fungsi tersebut harus sama dengan nol. Yang perlu diperhatikan di sini adalah adanya data X dan Y, sedangkan nilai yang tidak diketahui adalah a dan b, maka persamaan tersebut dicari turunan parsialnya terhadap a dan b, sehingga diperoleh persamaan sebagai berikut:

$$\frac{\partial \prod}{\partial a} = 2\sum_{i=1}^{n}[Y_i - (a + b\ X_i)] = 0$$

$$\frac{\partial \prod}{\partial b} = 2\sum_{i=1}^{n}X_i[Y_i - (a + b\ X_i)] = 0$$

Jika persamaan di atas diselesaikan, maka nantinya akan diperoleh:

$$\sum_{i=1}^{n} Y_i = a\sum_{i=1}^{n} a + b\ \sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_i Y_i = a\sum_{i=1}^{n} X_i\ + b\ \sum_{i=1}^{n} X_i{}^2$$

Rumus untuk memperoleh nilai a dan b menjadi:

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n \sum X^2 - (\sum X)^2}$$

$$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2}$$

## 2.2 PERHITUNGAN MANUAL

Tabel berikut adalah contoh data penjualan suatu produk (Y) dari waktu ke waktu (X). Untuk memudahkan perhitungan, penulis hanya menggunakan 10 data. Perlu dipahami bahwa di dalam sebuah penelitian biasanya data sampel-lah yang digunakan untuk membuat suatu persamaan, bukan data populasi. Kemudian persamaan yang terbentuk itu digunakan untuk memprediksi nilai populasi. Dengan demikian jumlah data yang digunakan kurang lebih akan memengaruhi persamaan yang dihasilkan yang pada gilirannya akan memengaruhi akurasi nilai prediksi. Agar persamaan yang dihasilkan itu lebih akurat, akan dibahas pula tentang uji asumsi klasik. Di samping itu, pembaca juga diharapkan memelajari bagaimana menentukan jumlah data yang digunakan pada penelitian sehingga hasilnya dapat dipertanggungjawabkan.

| X | Y |
|---:|---:|
| 1 | 60 |
| 2 | 90 |
| 3 | 75 |
| 4 | 100 |
| 5 | 125 |
| 6 | 115 |
| 7 | 130 |
| 8 | 125 |
| 9 | 160 |
| 10 | 150 |

Jika diasumsikan satuan waktu di sini adalah hari, maka dengan menggunakan analisis regresi sederhana dapat dilakukan estimasi penjualan pada hari-hari berikutnya. Tentu saja sebelum kita bisa melakukan estimasi, harus dicari terlebih dahulu persamaan regresinya.

Sebelum menentukan persamaan regresi, harus dihitung terlebih dahulu $\Sigma X$, $\Sigma Y$, $\Sigma X^2$, $\Sigma Y^2$, dan $\Sigma XY$. Untuk itu dibuatlah tabel berikut sehingga perhitungannya akan menjadi lebih mudah.

| X | Y | $X^2$ | $Y^2$ | XY |
|---|---|---|---|---|
| 1 | 60 | 1 | 3600 | 60 |
| 2 | 90 | 4 | 8100 | 180 |
| 3 | 75 | 9 | 5625 | 225 |
| 4 | 100 | 16 | 10000 | 400 |
| 5 | 125 | 25 | 15625 | 625 |
| 6 | 115 | 36 | 13225 | 690 |
| 7 | 130 | 49 | 16900 | 910 |
| 8 | 125 | 64 | 15625 | 1000 |
| 9 | 160 | 81 | 25600 | 1440 |
| 10 | 150 | 100 | 22500 | 1500 |
| 55 | 1130 | 385 | 136800 | 7030 |

Dengan bantuan tabel di atas, maka diperoleh $\Sigma X = 55$, $\Sigma Y = 1130$, $\Sigma X^2 = 385$, $\Sigma Y^2 = 136800$, dan $\Sigma XY = 7030$. Kemudian dihitung nilai b atau besarnya koefisien variabel X dengan menggunakan rumus berikut:

$b = [\ n\ \Sigma XY - \Sigma X\ \Sigma Y\ ] / [\ n\ \Sigma X^2 - (\Sigma X)^2\ ]$

$b = [\ (10 \times 7030) - (55 \times 1130)\ ] / [\ (10 \times 385\ ) - (55^2)\ ]$

$b = 9{,}88$

Setelah nilai b diperoleh, maka nilai a dapat dihitung dengan menggunakan rumus berikut:

$$a = \bar{Y} - b^{x}\bar{X}$$

di mana:

| | | |
|---|---|---|
| | $\bar{Y}$ | = rata-rata nilai Y = 1130 / 10 = 113 |
| | $\bar{X}$ | = rata-rata nilai X = 55 / 10 = 5,5 |
| Maka: | a | = 113 – (9,88 x 5,5) = 58,67 |

Diperoleh persamaan linier sebagai berikut: $\hat{Y}$ = 58,67 + 9,88 X

Apabila ditanyakan estimasi jumlah penjualan pada hari ke-20, maka hasilnya adalah:

$\hat{Y}$ = 58,67 + 9,88 X

$\hat{Y}$ = 58,67 + 9,88 (20)

$\hat{Y}$ = 256,27 ≈ 257

Analisis regresi biasanya bukan hanya sekedar menemukan persamaannya, tetapi juga menganalisis apakah pengaruh variabel bebasnya signifikan terhadap variabel terikatnya. Dengan demikian perlu dilakukan uji signifikansi, baik konstanta maupun koefisien dalam persamaan yang dihasilkan.

Informasi yang diperlukan untuk melakukan analisis signifikansi konstanta maupun koefisien regresi adalah kesalahan baku estimasi. Kesalahan baku estimasi dapat dihitung dengan menggunakan rumus:

$$se^2 = SSE / (n-k)$$

di mana

se = kesalahan baku estimasi

SSE = Sum of Square Error atau $\Sigma (Y - \hat{Y})^2$

n = jumlah data

k = jumlah variabel yang diamati

SEE dapat dihitung dengan bantuan tabel berikut:

| Y | Ŷ | Yi-Ŷi | (Yi-Ŷi)² |
|---|---|---|---|
| 60 | 68.55 | -8.55 | 73.02 |
| 90 | 78.42 | 11.58 | 134 |
| 75 | 88.3 | -13.3 | 176.97 |
| 100 | 98.18 | 1.82 | 3.31 |
| 125 | 108.06 | 16.94 | 286.94 |
| 115 | 117.94 | -2.94 | 8.64 |
| 130 | 127.82 | 2.18 | 4.76 |
| 125 | 137.7 | -12.7 | 161.21 |
| 160 | 147.58 | 12.42 | 154.36 |
| 150 | 157.45 | -7.45 | 55.57 |
| 1130 | | **SSE** | 1058.79 |

Apabila hasil data yang terdapat di dalam tabel di aplikasikan pada rumus di atas, maka kesalahan baku estimasi itu adalah:

$se^2$ = SSE / (n-k)

= 1058,79 / (10 – 2)

= 132,35

Dari kesalahan baku estimasi, bisa dihitung kesalahan baku konstanta dan kesalahan baku koefisien regresi dengan menggunakan rumus:

$S^2a = (se^2 * \Sigma X^2) / n \Sigma x^2$

$S^2b = se^2 / \Sigma x^2$

Untuk mempermudah perhitungan, maka dibuatkan tabel berikut ini:

| | | $Xi - \bar{X}$ | $Yi - \bar{Y}$ | | | |
|---|---|---|---|---|---|---|
| X | Y | x | y | $x^2$ | $y^2$ | xy |
| 1 | 60 | -4.5 | -53 | 20.25 | 2809 | 238.5 |
| 2 | 90 | -3.5 | -23 | 12.25 | 529 | 80.5 |
| 3 | 75 | -2.5 | -38 | 6.25 | 1444 | 95 |
| 4 | 100 | -1.5 | -13 | 2.25 | 169 | 19.5 |
| 5 | 125 | -0.5 | 12 | 0.25 | 144 | -6 |
| 6 | 115 | 0.5 | 2 | 0.25 | 4 | 1 |
| 7 | 130 | 1.5 | 17 | 2.25 | 289 | 25.5 |
| 8 | 125 | 2.5 | 12 | 6.25 | 144 | 30 |
| 9 | 160 | 3.5 | 47 | 12.25 | 2209 | 164.5 |
| 10 | 150 | 4.5 | 37 | 20.25 | 1369 | 166.5 |
| 55 | 1130 | 0 | 0 | 82.5 | 9110 | 815 |

Perhatikanlah penggunaan huruf besar dan huruf kecil. Dari tabel tersebut terlihat bahwa $x = Xi - \bar{X}$ dan $y = Yi - \bar{Y}$. Sehingga diperoleh $S^2a$ dan $S^2b$ adalah sebagai berikut:

$S^2a$ = $(se^2 * \Sigma X^2) / n\ \Sigma x^2$

= (132,35 * 55) / 10 (82,5)

= 61,76

Sa = 7,86

$S^2b$ = $se^2 / \Sigma x^2$

= 132,35 / 82,5

= 1,6

Sb = 1,27

Dari kesalahan baku konstanta dan koefisien regresi, kemudian dicari nilai $t_{hitung}$ untuk kosntanta dan koefisien regresi dengan menggunakan rumus berikut ini:

$t_{hitung}$ (a) = a / Sa

= 58,67 / 7,86

= 7,46

$t_{hitung}$ (b) = b / Sb

= 9,88 / 1,27

= 7,8

Syarat menerima H0, yang menyatakan menerima konstanta atau koefisien regresi tidak signifikan, adalah: $-t_{tabel} \leq t_{hitung} \leq t_{tabel}$. Untuk taraf signifikansi 0,05 dan derajat kebebasan (n-k) = 10 – 2 = 8, diperoleh $t_{tabel}$ sebesar 2,306 (uji 2 pihak atau 2-*tail*). Karena $t_{hitung}$ untuk konstanta dan koefisien regresi tersebut berada di luar area yang menyatakan menerima H0, maka konstanta dan koefisien regresi tadi terbukti signifikan. Karena koefisien regresi signifikan, maka bisa dikatakan pengaruh variabel bebas terhadap variabel terikat terbukti signifikan.

## 2.3 PERHITUNGAN SPSS

Untuk mendapatkan persamaan regresi seperti yang ditampilkan pada pembahasan sebelumnya, lakukan langkah-langkah berikut:

1. Masukkan data dan sesuaikan tipe datanya. Silakan definisikan tipe data pada bagian Variabel View,
2. Pilih *Analyze-Regression-Linear* pada menu program SPSS,
3. Kemudian masukkan variabel terikat (*dependent*) dan variabel bebas (*independent*),
4. Setelah itu tekan tombol "OK".

Data yang telah dimasukkan ke dalam program SPSS itu akan tampil seperti berikut ini:

Simple Regression.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

10 :

| | X | Y | var | var | var | var |
|---|---|---|---|---|---|---|
| 1 | 1 | 60 | | | | |
| 2 | 2 | 90 | | | | |
| 3 | 3 | 75 | | | | |
| 4 | 4 | 100 | | | | |
| 5 | 5 | 125 | | | | |
| 6 | 6 | 115 | | | | |
| 7 | 7 | 130 | | | | |
| 8 | 8 | 125 | | | | |
| 9 | 9 | 160 | | | | |
| 10 | 10 | 150 | | | | |

Hasil keluaran (*output*) dari program SPSS adalah seperti gambar di bawah ini.

**Variables Entered/Removed**[b]

| Model | Variables Entered | Variables Removed | Method |
|---|---|---|---|
| 1 | X[a] | . | Enter |

a. All requested variables entered.

b. Dependent Variable: Y

**Coefficients**[a]

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 58.667 | 7.859 | | 7.465 | .000 |
| | X | 9.879 | 1.267 | .940 | 7.800 | .000 |

a. Dependent Variable: Y

Nilai koefisien regresi dan konstanta yang dihasilkan oleh program SPSS dapat dilihat pada kolom *Unstandardized Coefficients* (kolom B). Berdasarkan tabel tersebut, maka persamaan liniernya adalah:

$\hat{Y} = 58{,}667 + 9{,}879\ X$.

Hasil perhitungan manualnya adalah:

$\hat{Y} = 58{,}67 + 9{,}88\ X$.

Perhatikan juga bahwa nilai $t_{hitung}$ yang dihasilkan oleh program SPSS sama dengan perhitungan manualnya yaitu 7,465 dan 7,800.

# BAB III
# ANALISIS REGRESI BERGANDA

## 3.1 KONSEP DASAR

Yang dimaksud dengan analisis regresi berganda di sini adalah analisis yang dilakukan terhadap satu variabel terikat dan dua atau lebih variabel bebas. Persamaan yang dihasilkan adalah persamaan dengan bentuk $\hat{Y} = f(X1, X2, ..., Xn)$. Jadi perbedaan regresi sederhana dan regresi berganda hanya terletak pada jumlah variabel bebas yang digunakannya. Jika ada dua variabel bebas yang digunakan ($X_1$ dan $X_2$), maka persamaannya dapat ditulis menjadi $\hat{Y} = a + b X_1 + c X_2$. Lambang $\hat{Y}$ (baca: *y topi*) digunakan untuk menyatakan bahwa data yang diperoleh dari persamaan regresi adalah data prediksi. Sedangkan data aktual untuk variabel terikat ditulis dengan lambang Y.

Pada metode OLS, nilai a, b dan c dapat dihitung dengan meminimalkan fungsi berikut:

$$\sum_{i=1}^{n} e_i^{\,2} = \sum_{i=1}^{n} [Y_i - f(X_{1i}, X_{2i})]^2 = minimum$$